Assessment Instrument Description: **aimsweb**

Element	Element Description	
Instrument Name	Name of specific instrument (more than vendor name).	**aimsweb**
Vendor	Name of the company or organization that produces the instrument.	NCS Pearson, Inc.
Purpose (Intended Use)	The described purpose and appropriate uses of the instrument.	**aimsweb** is a benchmark and progress monitoring system based on direct, frequent and continuous student assessment using brief, accurate measures of reading, math, spelling, and writing. **aimsweb** is the most comprehensive K-12 assessment system that supports Response to Intervention (RTI) and tiered instruction. The results are reported to students, parents, teachers and administrators via a web-based data management and reporting system to determine response to intervention. Results allow educators to effectively allocate limited resources by efficiently identifying students at risk, monitoring progress, and differentiating instruction.
Types of Instruments	Interim, Summative, Diagnostic	Interim, universal screening, progress monitoring
Population	Who (which students) could be assessed using the instrument.	Aimsweb assesses students in Kindergarten through Grade 12. Tier 1 - Assess all students three times per year for universal screening (early identification), general education progress monitoring, and AYP accountability. Tier 2 - Assess and monitor at-risk students and the effectiveness of instructional changes. Tier 3 - Write individualized annual goals and monitor progress more frequently for those who need intensive instructional services.
When? How frequently?	How frequently the instrument can be administered in a school year, and recommended or required administration windows.	Benchmarking is designed to inform instruction to improve achievement. Benchmarks are established three times per year for all students, based on established school and district windows. The school district determines the screening periods, depending on the local school schedule, but two-week windows generally are recommended during • The first four weeks of school • The midpoint of the year (around the eighth week of school)

		• The last four weeks of school.
		Strategic Monitor provides schools with the option to increase assessment frequency for students who have been identified as "struggling," or minimally at-risk in the Benchmark process – or for all students, if desired. Increasing assessment frequency provides more opportunity to evaluate the effectiveness of instructional changes and to verify struggling student achievement levels or to confirm there is no degeneration of progress in minimally at-risk students.
		Progress Monitor is a continuous assessment and improvement system designed specifically for frequent assessment and monitoring of at-risk students, including those receiving Title I services, or those identified with a learning disability or other special needs.
Content Area (s)	Content area or areas being assessed.	**aimsweb** offers assessments in each of the following areas: • Early Literacy (English and Spanish) • Reading (English and Spanish) • Written Expression • Spelling
Learning Objectives	Specific learning objectives being assessed, at as detailed a level as is provided. This may be "topics" or categories or may be actual learning objective statements.	(see table below)

Early Literacy

Letter Naming Fluency	Student says the names of visually presented letters; number of corrects are scored
Letter Sound Fluency	Student says the sounds of visually presented letter; number of corrects are scored
Phoneme Segmentation	Student hears orally-presented words and says their component phonemes. The score is the number of segments said correctly in one minute.
Nonsense Word Fluency	Student says the sounds of visually presented non-real words; number of corrects are scored

MIDE (Spanish Early Literacy)

Letter Naming Fluency	Student says the names of visually presented letters; number of corrects are scored
Letter Sound Fluency	Student says the sounds of visually presented letter; number of corrects are scored
Syllable Segmentation Fluency	Student identifies the specific syllables in orally presented words; number of corrects

Syllable Reading Fluency	Student reads CV syllables presented visually; number of corrects are scored
Syllable / Word Spelling	Student writes CV syllables (K) or primary vocabulary words (Grade 1) as prescribed orally; correct letter sequences and number of correctly spelled syllables/words are scored
Reading (English)	
Reading-CBM	Oral reading of a grade-appropriate narrative fiction passage of 150 to 400 words. For screening, three probes are administered. The final score is the median score on the three probes. The score for a probe is the number of words read correctly; the number of errors is also recorded. Development: The oral reading passages are derived from narrative fiction passages carefully written and tested with students to ensure that the passages within each grade level are similar in difficulty.
Reading (Spanish)	
Spanish Reading-CBM	Student reads passage aloud; number of words read correctly are scored.
Spanish MAZE	A multiple-choice cloze task that students complete while reading silently. The first sentence of the 150-400 word passage is left intact. Thereafter, every 7th word is replaced with three words inside parentheses, from which the student selects the correct word. The # of corrects are scored. Development: The passages are the same as those used for R-CBM.
Written Expression	
Written Expression CBM	Student hears a "story starter" and has four minutes to write a narrative (one minute of mental preparation and three minutes of writing).

		Three scores may be obtained: Total Words Written; number of Correct Writing Sequences (adjacent words that are correct in meaning, syntax, grammar, spelling, and mechanics); and Words Spelled Correctly.". 1-8
	Spelling	
	Spelling CBM	The student writes dictated words presented at 10-second (grades 1 and 2) or 7-second (grades 3-8) intervals. Two score may be obtained: Word Spelled Correctly, and Correct Letter Sequences
Individual Metrics	The scores provided at the individual (student) level.	Student Level Reports include: **Benchmark** - For benchmark screening data, individual student scores for fall, winter, and spring are displayed in conjunction with the score distribution for a selected reference group, which is represented using a "box and whiskers" format. Screening results can be interpreted through both criterion-referenced and norm-referenced methods. • A criterion-referenced interpretation compares a student's scores with designated scores that indicate a good likelihood of academic success. The criterion scores may be based on expert judgment or on an empirical demonstration of the relationship between the screening score and a positive outcome. • A norm-referenced interpretation compares a student's score with the scores of other students in a local or national reference group of students in the same grade tested on the same measure at the same time of year. **aimsweb** provides norm-referenced information in the form of a percentile, that is, the percentage of students in the reference group who scored below a particular score. On this scale, a score at the 50th percentile is average (higher than half the students in the norm sample), the 10th percentile is very low, and the 90th percentile is very high. Both national and local (school and/or district) percentile norms are provided by **aimsweb**. **Progress Monitor** - Students who are identified as at-risk in the screening process are progress monitored. Educators carry out interventions, establish goals, and monitor progress to measure the effectiveness of instructional changes and progress towards the goal. Progress monitoring in **aimsweb** supports reassessing the student as needed, up to once per week. **aimsweb** progress monitoring assessments are sensitive to small changes in performance allowing educators to make decisions on intervention changes

		in a timely manner. If the student's rate of improvement is not sufficient to meet the goal, the educator introduces a new intervention—which is noted on the chart as a vertical line and documented in the report narrative—and a new rate of improvement is calculated for that intervention. Individual student progress monitor reports offer significant detail for educators. The table above (in our response to Learning Objectives) provides some information on scores for each assessment. The Scoring and Administration Manuals provide detailed more detailed information.

Individual Comparison Points (cut scores)	Information provided regarding how good is good enough performance on the instrument. Comparison information should be available for every individual metric. This may be performance level ratings with specific cut scores.	Student results are provided in a score for each measure by screening period (F, W, S). Information provided at the individual student level for the measures includes
		• Raw scores
		• National Percentile
		• Lexile (for certain measures only)
		• Rate of Improvement
		    ○ Student Rate of Improvement
		    ○ National Rate of Improvement
		    ○ Student Growth percentile
		For progress monitoring on individual measures the following are reported:
		• Raw score
		• Errors
		• Goal rate of improvement
		• Trend rate of improvement
		• Aimline (a line connecting the baseline score to the goal score)
		• Trendline
		• Student's likelihood of meeting the performance goal by the goal date.
		
		Scores obtained during universal screening can be interpreted using national or local (school and/or district) percentile norms for that grade and screening period (fall, winter, or spring). The national percentile norms are based on large, representative samples of students. The AIMSweb system is able to provide local norms once a sufficient amount of data has been entered. When generating group or individual reports, you may select to show national norms, local norms, or both. Some of the reports for R–CBM and Reading Maze also provide Lexile® measures that correspond to the AIMSweb raw scores. The Lexile measure indicates text difficulty and reading ability. It enables educators to choose books or other reading materials appropriate for a student—neither too easy to be

challenging, nor too difficult to be comprehensible. For more detailed information, please refer to the AIMSweb Introductory guide.

Aggregate Metrics	Scores provided at the group level. The group could be a grade level, school, district, or disaggregated groups (e.g. race/ethnicity, gender, IEP status, FRL status) Specify the group(s) and the score(s) provided.	Here is a list of the reporting screens available in the **aimsweb**. • Student Individual Benchmark (several iterations) • Student Individual Progress Monitoring • Teacher-level reports such as the Summary Report • Summary Reports at the class, grade, school, district, or state levels • Subgroup Reports with disaggregated data • RTI Tier Reports • ELL Profile Reports • Student Growth Percentile Reports • Longitudinal Data Reports
Aggregate Comparison Points (cut scores) Vendor	Information provided regarding how good is good enough performance at the group level.	**aimsweb** recommends using the 15th and 45th national percentiles as follows: • Not On Track: ≤15th national percentile • Further assessment may be needed: 16th–45th national percentile • On Track: >45th national percentile
Comparison Points (CDE)	CDE cut scores for requests to reconsider.	Based on the body of research on R-CBM and M-CAP, **aimsweb** researchers were able to select default cut scores for Reading Maze, M-COMP, Written Expression, and Spelling. When the researchers reviewed the percentile values of the R-CBM and M-CAP cut scores on the new National Norms, they were struck by the high level of consistency of the percentiles across grades, benchmark periods, and measures. For both measures, the 80% Success Probability score was consistently close to the **45th percentile** and the 50% Success Probability score was consistently near the **15th percentile**. For this reason, **aimsweb** concluded that it would be reasonable to use those percentile values to set default cut scores for other measures of reading, language arts, and math. For Maze, M-COMP, Written Expression, and Spelling, the default cut scores should not be interpreted as predictors of state test success, because they are not based on direct empirical evidence involving scores on those measures. The rationale for these cut scores is that if the lowest-scoring **15%** of the national student population has consistently been found to be at severe risk in reading and math, and the lowest-scoring **45%** at moderate

		risk, then it is reasonable to use those percentages as a guide to the number of students who should be identified as at-risk when using other measures. This method has the benefit of being grounded in empirical research, rather than using theoretical or arbitrary percentile cutoffs.
		aimsweb Default Cut Scores for TEL and TEN The default cut scores for the Test of Early Literacy (TEL) and Test of Early Numeracy (TEN) measures were established in a similar way, but using a different criterion for success. Silberglitt (2001) did a study of the relationship of **aimsweb** TEL scores to reading success in Grade 2 as measured by adequate R-CBM performance. He identified the raw scores on Letter Sound Fluency (by grade and period) that predicted success, and these are consistently close to the 35th percentile on the new **aimsweb** National Norms. That percentile value is used for the higher cut score on the TEL and TEN measures; the lower cut score is set at the 15th percentile, consistent with its location for the other **aimsweb** measures.
Alignment	Information provided by the vendor about alignment of this instrument to other instruments, standards, etc.	**aimsweb** measures have a crosswalk to the Common Core State Standards (CCSS) and to the learning standards for many individual states, including Colorado.
Data Reports	Description of data reports that are provided/available at the individual and aggregate level(s).	Here is a list of the reporting screens available in the **aimsweb**. • Student Individual Benchmark (several iterations) • Student Individual Progress Monitoring • Teacher-level reports such as the Summary Report • Summary Reports at the class, grade, school, district, or state levels • Subgroup Reports with disaggregated data • RTI Tier Reports • ELL Profile Reports • Student Growth Percentile Reports • Longitudinal Data Reports
Technical Quality		The technical properties that are important for a given assessment depend on how educators will use and interpret the assessment. For assessment systems such as **aimsweb** that employ general outcome measures (GOMs) for universal screening

and progress monitoring, highly important technical properties include the following.

- **Equivalence** of probe difficulty, within each grade level. A student's expected score should be the same regardless of which grade-level probe is administered. This feature is important for Response to Intervention (RTI) when progress monitoring so that the trend of scores across time will not be unduly influenced by how the difficulty of probes varies.

- **Reliability** of probe scores. Reliability refers to the consistency, or repeatability, of scores. It may also be thought of as the degree to which scores are free of measurement error. Two types of reliability are particularly relevant to the **aims**web measures:

  o **Alternate-form reliability:** The agreement between scores on alternate forms (probes) administered relatively close together in time. This type of reliability indicates how free the score is from changes due to day-to-day fluctuations and from differences in the specific content of the probes. (Note that if a measure has only one probe, such as Oral Counting, then **test–retest reliability** indicates consistency over time.)

  o **Interrater reliability:** The agreement among scores calculated by independent raters. This type of reliability is especially relevant when scoring requires judgment.

Internal-consistency reliability (such as split-half or coefficient alpha) is generally not suitable for **aims**web measures because they assess speed as well as accuracy, and most students do not reach the end of the probe within the time limit.

- **Validity** of probe scores. Validity refers to the accuracy of inferences made from scores, such as, "This student is better at math calculation than about 40% of his or her peers," or, "This student has a high probability of passing the state reading test." The inferences most often made from **aims**web scores are based on assumptions about the probes' content and on the relationship of probe scores with external criteria.

  o **Content validity:** The degree to which the test score measures the designated knowledge/skill domain. Evidence of content validity is

typically obtained by comparing item content with a curriculum or with how well the tasks address the purported purpose of the measure.

- o **Criterion validity:** The relationship between test scores and a criterion, such as success in an educational program or scores on other tests. Two subtypes of criterion validity are often differentiated according to the amount of time between the **aims**web administration and the occurrence of the criterion:

  - **Concurrent validity** is the correlation of probe scores with criterion data (e.g., another test) collected at the same time.

  - **Predictive validity** is the correlation of probe scores with a future criterion.

- o **Classification accuracy:** An alternative means of expressing criterion validity that is appropriate when there is interest in predicting a dichotomous criterion (e.g., passing or not passing a state test). A cut score on the predictor test (i.e., the **aims**web measure) is chosen such that those who score at or above the cut score are considered likely to pass the criterion, while those who score below the cut score are likely to fail. A classification-accuracy analysis indicates how frequently these expectations prove correct, and the results are reported in a variety of statistics, including:

  - **Sensitivity,** which refers to the proportion of those students who actually fail the criterion who score below the **aims**web cut score.

  - **Specificity,** which refers to the proportion of those students who actually pass the criterion who score above the **aims**web cut score.

  - **Area under the curve,** a global indicator of the degree to which the **aims**web measure correctly predicts the outcome on the criterion.

In addition to these technical characteristics of scores obtained at a single point in time, there are technical properties related to a student's rate of improvement

(ROI, or slope) during progress monitoring. The **standard error of measurement (*SEM*) of the ROI** is important to consider when making a decision about whether a student's progress is sufficient to reach the established goal. This *SEM* is a function of the number and variability of the student's scores and their time span. For example, scores that closely follow a straight line across a wide span of time produce a small ROI *SEM*. The average size of the ROI *SEM* relative to the total amount of variability of the ROIs in a population of students indicates the **reliability of the ROI.** Furthermore, the **criterion validity of the ROI** can be evaluated by investigating whether students who show a faster rate of progress on an **aims**web measure are likely to score higher on a criterion (after controlling for the initial score level).